

## Vers Un Système De Traduction Automatique En Ligne Des Documents Amazighes Base Sur Les Graphes UNL

Ali Rachidi

*Ecole nationale de Commerce et de Gestion, B. P. 37/S Hay Salam, IRF - SIC, Faculté des sciences, Université Ibn Zohr, Agadir, Maroc, rachidi.ali@caramail.com*

Driss Mammass

*IRF - SIC, Faculté des Sciences, B.P. 8106, Hay Dakhla, Université Ibn Zohr, Agadir, Maroc, driss\_mammass@yahoo.fr*

### Résumé

---

L'intégration des technologies de l'information et de communication (TIC) à l'apprentissage de la langue Amazighe est absolument nécessaire pour que l'Amazighe ait droit de cité plein et entier sur le Web et dans le monde informatisé. Dans ce contexte, il est de plus en plus nécessaire de créer des documents multilingues qui intègrent l'Amazighe. La coédition d'un texte en langue naturelle et de sa représentation dans une forme interlingua semble le moyen le meilleur et le plus simple de partager la révision du texte vers plusieurs langues et de répercuter du texte et les modifications de L0 vers l'IL, et de régénérer vers L1,... Ln depuis l'IL.

Il faudra cependant permettre des améliorations manuelles, car la forme interlingua ne sera pas toujours présente, ou pas assez améliorable par défaut d'expressivité, et les générateurs ne seront jamais parfaits. Pour plusieurs raisons, les graphes UNL (Universal Networking language) sont les meilleurs candidats. L'idée actuelle est de faire de la traduction manuelle collaborative sur le Web à l'aide d'une mémoire de polyphrases multilingues (MPM) outils construit par l'équipe de M. C. Boitet (GETA, CLIPS, IMAG à Grenoble, France), puis d'intégrer le résultat (les phrases en Amazighe) dans le document en UNL-XML. Enfin, il faudra construire un déconvertisseur UNL-Amazighe et un enconvertisseur en se basant sur les bases de connaissances construites sur la base des objets typés.

### Abstract

---

The use of IT tools with Tamazight Berber is an absolute requisite for giving this language full citizenship on the Web in particular and in the digital world in general. Thus, the need to create Tamazight digital documents is becoming increasingly urgent. Granted, Unicode now includes the full Tifinagh character set, but the question that remains is how to implement information interchange between Tamazight and the languages of the Web. We contend that the best and simplest way to share concurrent revisions of the same text in multiple languages is to coedit text in natural language and then to render it into Interlingua (IL) for dissemination. This method allows the participants to (i) translate the text, with modifications if any, from L0 into IL, and subsequently to (ii) regenerate the text in L1 ... Ln starting from the same IL representation.

Generators will never be perfect. Therefore, manual editing should always be allowed, because the automatically generated IL form may be irremediably inexpressive, or may altogether be unavailable due to lack of relevant data in the knowledge base. Since Universal Networking Language (UNL) graphs seem to be the best tool for the job, human participants should use a UNL editor for manual translation and revising. We propose that the collaborative manual translations be done on the Web with the help of the bank of multilingual utterances compiled by C. Boitet's team (GETA, CLIPS, IMAG à Grenoble, France). The resulting Tamazight utterances should then be integrated in a UNLXML document. At a later stage it will be necessary to build a UNLTamazight deconverter and a TamazightUNL enconverter, relying on the knowledge base built on top of the objects that have been tagged so far.

### Mots-clés

---

communication multilingue, graphe UNL, représentation interlingua, tifinaghe

### Keywords

---

multilingual communication, UNL graph, interlingua representation, tifinagh

## 1. Introduction

Ce papier s'inscrit dans un large mouvement international qui vise à ce que chaque peuple puisse disposer de tous les moyens pour communiquer dans sa langue. A l'époque, affirmer ou défendre une langue passait par d'autres moyens : fixer une orthographe, construire des dictionnaires monolingues ou bilingues, recueillir des traditions orales ou encore élaborer des polices d'imprimeur.

Aujourd'hui, le développement des ordinateurs personnels et des réseaux fait de l'informatique un moyen pour écrire et communiquer au même titre que le papier l'est depuis *Cai Lun* et l'imprimerie depuis *Gutenberg*. Mais les langues ne sont pas égales devant le processus d'informatisation et les populations parlant des langues mal dotées ont un accès limité à ces nouveaux moyens, limitation pouvant aller d'une simple gêne à une incapacité totale. L'**Amazighe** fait partie de ces langues peu dotées informatiquement. Par conséquent, des recherches scientifiques et linguistiques sont lancées dans ce sens pour améliorer la situation actuelle (Rachidi et Mammas, 2005a) (Rachidi et Mammas, 2005b) (Rachidi et Mammas, 2005c) (Boukous et al., 2003). Parmi les volets, qui occupent une position prioritaire, est de concevoir et réaliser des documents multilingues qui intègrent la langue Amazighe. En effet, et dans la pratique courante, un document multilingue consiste en plusieurs fichiers monolingues parallèles, qui peuvent être de la documentation technique aussi bien que des fichiers d'aide, des fichiers de messages, ou simplement de l'information thématique mise sur la toile et destinée à une audience multilingue (médecine, cuisine, chanson, traditions...).

La tâche est difficile, même pour un document géré de façon centralisée (Al Assimi, 2000) (Al Assimi et Boitet, 2001). D'habitude, on le crée d'abord dans une unique langue source, puis on le traduit dans plusieurs langues cibles (au moins une trentaine dans le cas de firmes comme IBM, HP ou Microsoft). Il faut avoir un moyen de garder trace des modifications, qui peuvent parfois être faites en différents lieux sur des versions en différentes langues. Dans ce cas, des traducteurs humains doivent retraduire les segments nouveaux ou modifiés, ou les réviser s'ils sont traduits par un système de TA de qualité. Ce que nous souhaiterons, c'est faire en sorte que le travail de révision puisse être partagé entre les langues, quels que soient le domaine et le contexte.

Il est clairement impossible de refléter les changements sur un fichier en langue L0 dans les fichiers en langues L1,... Ln automatiquement et fidèlement, sans une structure intermédiaire pour faire le pont, car il faudrait au moins un aligneur parfait à granularité très fine dans le cas simple d'un changement d'article ou de nom.

L'approche la meilleure et la plus simple nous semble être d'utiliser un interlingua formel IL et de :

- répercuter les modifications de L0 vers l'IL,
- régénérer vers L1,... Ln depuis l'IL.

Il faudra cependant permettre des améliorations manuelles, car la forme interlingua ne sera pas toujours présente, ou pas assez améliorable par défaut d'expressivité, et les générateurs ne seront jamais parfaits.

On choisit UNL (Blanc, 2001) (Boguslavsky et al., 2000) (Sérasset et Boitet, 99) (Sérasset et Boitet, 2000) comme interlingua pour différentes raisons :

- il est spécialement conçu pour le traitement linguistique et sémantique par ordinateur ;
- il a été dérivé avec beaucoup d'améliorations du langage pivot de H. Uchida utilisé dans ATLAS-II de Fujitsu (Uchida, 89), toujours évalué comme le système de TA anglais-japonais de meilleure qualité, avec une très grande couverture (586.000 entrées par langue) ;
- les participants du projet UNL<sup>1</sup> ont construit des "déconvertisseurs" d'UNL vers environ 12 langues, parmi elles, on cite l'arabe, l'indonésien, l'italien, le français, le russe, l'espagnol et le thaï;

<sup>1</sup> <http://unl.ias.unu.edu>

- bien qu'ils soient de nature formelle, les graphes UNL (voir ci-dessous) sont assez simples à comprendre avec peu de formation et peuvent être présentés de façon localisée à des utilisateurs "naïfs" en traduisant les symboles (relations sémantiques, attributs) et les lexèmes du langage UNL par des symboles et des lexèmes de leur langue ;
- le projet UNL a défini un format "UNL-html" intégré à html pour des fichiers contenant un document multilingue complet aligné au niveau des énoncés, et a produit un "visualiseur" qui transforme un fichier dans ce format en autant de fichiers html que de langues, et les envoie à n'importe quel navigateur Web.

La suite est organisée comme suit. D'abord, on présentera l'UNL comme langage de représentation unifié. Ensuite, on exposera le format UNL-HTML comme l'un des formats des documents multilingues stockés sur un serveur sous forme d'une collection de fichiers multilingues. Enfin, on essaie de donner une idée sur ce qui sera le déconvertisseur Unl-Amazighe et l'enconvertisseur associé en passant sur un certain nombre de concepts et définitions.

## 2. Langage UNL (Universal Networking Language)

L'UNL est l'acronyme de ' Universal Networking Language'. C'est un langage artificiel qui réplique, dans le Web, les fonctions des langages naturels dans la communication humaine. Il permet aussi aux ordinateurs du monde entier de inter communiquer, et enfin, avoir une infrastructure linguistique pour distribuer, recevoir et comprendre les informations multilingues. Par conséquent, il permet aux différentes populations du monde d'exprimer toutes les connaissances communiquées par les langages naturels.

L'UNL représente trois éléments caractéristiques: un projet international, un langage artificiel et un format de document multilingue.

### 2. 1. Le projet

Le projet UNLP consiste à mettre en place un système de traduction automatique universel. Il est défini comme un projet de "métalangage numérique" pour l'encodage, le stockage, la recherche et la communication d'informations multilingues indépendamment d'une langue source - et donc d'un système de pensée.

Il est mené sous l'égide de l'Université des Nations Unies (UNU, Tokyo). Plusieurs équipes réparties dans le monde travaillent sur ce projet (Japon, Inde, France, Espagne, Italie, Chine, Russie...). L'intérêt de cette collaboration est de travailler avec les mêmes règles et codes et d'obtenir ainsi un résultat vraiment universel, c'est-à-dire compréhensible par toutes les cultures.

### 2. 2. Le langage artificiel

L'UNL est un langage compris par les ordinateurs. Il est universel, c'est à dire que toute langue naturelle (parlée par des êtres humains) peut être traduite en UNL. De même, l'UNL peut être traduit en toute langue naturelle. Pour passer d'une langue naturelle à l'UNL, on utilise un enconvertisseur et pour passer de l'UNL à une langue naturelle, on utilise un deconvertisseur. La figure 1 démontre ce principe.

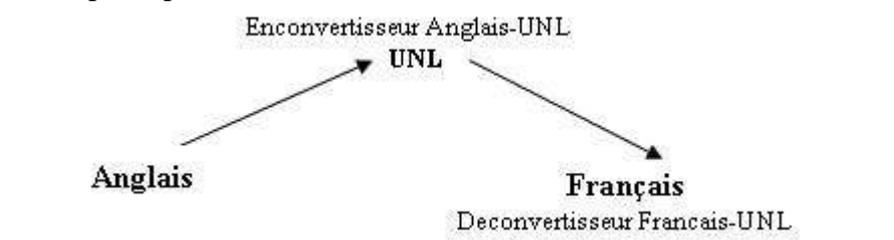


Figure 1. Principe de base de la traduction par UNL

Grâce à un enconvertisseur anglais-UNL et un deconvertisseur UNL-Français, on peut traduire automatiquement de l'Anglais au Français (cf. figure 2).



Figure 2 : principe de traduction

L'ajout d'un seul enconvertisseur Amazighe-UNL permet de passer de l'Amazighe au français (cf. figure 3). Ainsi, une équipe Amazigheienne n'a pas besoin de connaître le français pour faire progresser la traduction Amazighe-Français.



Figure 3. Principe d'ajout d'un enconvertisseur Amazighe-UNL

Enfin, deux équipes, française et anglaise, peuvent faire avancer les traductions Français-Anglais et indirectement Amazighe-Anglais. Et si on a le deconvertisseur UNL-Amazighe, la communication entre l'Amazighe, le Français et l'Anglais sera complète. Tout travail national aide à la communauté internationale.

### 2. 3. Le format de documents multilingues

Un même document UNL contient les textes en plusieurs langues ainsi qu'en UNL. Cependant, l'utilisateur ne verra que les langues le concernant, par exemple la langue source et la langue traduite. Le document est balisé, c'est-à-dire que des groupes de mots sont entourés de chaînes de caractères représentant un code compréhensible par les ordinateurs (comme <html> par exemple). On traitera UNL-HTML en détail dans la prochaine section.

## 3. Principes

L'UNL est composé de mots (Universal Words ou UW), de relations définies entre les mots et d'attributs. Les UW représentent le vocabulaire. Les relations entre UWs et les attributs représentent la syntaxe et la sémantique (donner un sens).

Comme l'anglais est connu par tous les développeurs informatiques, les UW sont définies par des termes ou des abréviations anglaises. Cependant, ceci n'est qu'une règle de commodité, les UWs pourraient être définies numériquement.

Les développeurs UNL francophones ont féminisé UW, car le terme français le plus voisin est "unité lexicale".

### 3. 1. Unités lexicales (UWs)

Les UWs représentent donc le vocabulaire de l'UNL. En d'autres termes, chaque UW représente un concept ou un ensemble de concepts, un concept correspondant à une idée précise. Une UW est formée d'un mot anglais suivi d'une liste de contraintes, qui précise le sens de l'UW.

La syntaxe d'une UW est :

<UW> := <Entrée> [liste des contraintes>] (entrée est un mot anglais)

Exemples :

Le mot Français « danse » existe au moins sous deux concepts, "je danse" ou "j'aime cette danse".

Dans le premier cas, il sera défini par `dance(icl>do)` ; c'est-à-dire comme un verbe d'action, défini par l'attribut "do" (faire en anglais).

Dans le second cas, il est défini par `dance(icl>thing)` ; c'est-à-dire comme un nom commun définissant un objet.

Le mot Amazighe «  $\text{ⵜⵓⵍⵓⵏ}$  » (TAWADA) sera défini par `walking(icl>thing)` et pas seulement « walking » puisque le mot désigne plusieurs choses.

En augmentant le nombre d'attributs, on définit très précisément le sens de chaque UW. Exemple, le terme "marquer" pour marquer un but dans un match de football est défini par `score(icl>event,agt>human, fld>sport)`. Ce mot est un verbe d'événement (`icl>event`), l'objet qui agit, l'agent, est un être humain (`agt>human`), le champ concerné est le sport (`fld>sport`). On remarque que dans la définition, le sens du mot est donné ainsi que les autres types de mot avec lesquels il peut interagir. Ceci pour éviter des phrases du genre : "Le chien a marqué un but de la main."

Il existe six types de UWs :

- UW de base est une entrée (mot en anglais) sans contraintes. Elle signifie tous les concepts correspondants en anglais. Elle est utilisée pour structurer la base de connaissances et comme moyens pour établir les correspondances entre les mots de différentes langues une fois que des correspondances spécifiques n'existent pas. Exemple : go, take, state, ...
- UW restreinte est une entrée suivie d'une liste de contraintes. Exemple: 1) `state(icl>express)`, 2) `state(icl>country)`, 3) `state(icl>government)`, etc.
- UW importée est une entrée qui n'est pas un mot anglais mais importée pour pouvoir exprimer des phrases en d'autres langues. Exemples : `samba(icl>dance)` ou  $\text{ⴰⵏⵉⵎⵉⵔⵉⵏ}$  : `ahidous(icl>dance)`.
- UW temporaire est une entrée qu'on ne doit pas intégrer dans la base de connaissances ou dans le vocabulaire des UWs. Exemple : formule, nom propre, etc.
- UW composée est une entrée temporaire, mais elle est un ensemble de relations binaires regroupées pour exprimer un concept complexe. Exemple : « women who wear big hats in movie theaters (`icl>person`) » peut être décomposée en plusieurs UWs.
- UW nulle est une chaîne de caractères vides pour exprimer l'absence d'un argument nécessaire d'une relation. Elle peut être considérée comme temporaire. Exemple : « Med was killed ». le sujet est absent. Pour exprimer l'idée, on utilise `agt(kill(agt>thing, obj>thing).@entry .@past, ' ')` ou `obj(kill(agt>thing,obj>thing).@entry.@past, Med.@topic)`

### 3. 2. Relations entre les UWs

Les relations entre les UWs sont définies par des chaînes de trois caractères ou moins. Ce sont des relations binaires (entre deux UWs) et elles permettent de préciser quelle UW peut être utilisée avec quelle UW. L'UNL traite les textes phrase par phrase; ainsi les relations entre les UWs sont définies à chaque phrase. Une phrase est représentée sous la forme d'un graphe ou d'une liste.

Exemple :

la phrase « I ran in the park yesterday » sera représentée par des relations (cf. figure 4.1) ou par des graphes (cf. figure 4.2).

Agt(run(icl>do).@entry.@past,i(icl>person))  
 Plc(run(icl>do).@entry.@past,park(icl>place).@def)  
 Tim(run(icl>do).@entry.@past,yesterday)

Figure 4.1. Relations entre les UWs

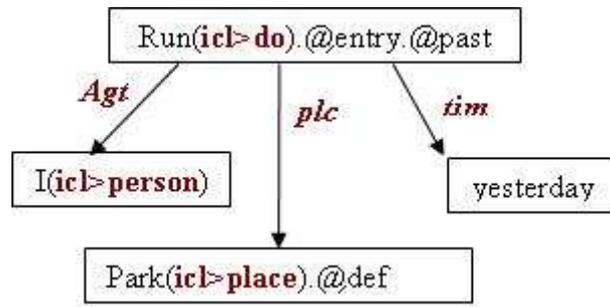


Figure 4.2. Graphe des relations entre des UWs

Une phrase UNL est centrée autour du verbe principal de la phrase, ici "run". Les relations entre les mots sont ensuite définies à partir de ce verbe.

Les ensembles en gras sont inclus dans la définition des UWs, indépendamment de la phrase (exemple : "run(icl>do)", le mot "run" est défini comme un verbe d'action par le code "do" ("faire" en anglais)). Les UWs utilisées ont des attributs :

- @entry signifie que l'UW est une entrée ou une UW principale de la phrase ;
- @past signifie que l'UW est au passé ;
- @def signifie que l'UW est défini (the park).

Les éléments en italique représentent les relations entre les mots dans la phrase (Agt, tim, plc)

Ce graphe sera attaché à une base de connaissances UNL (UNL KB) qui contient les définitions et les différentes catégories de mots d'une langue naturelle. Par exemple, dans l'exemple précédent, on utilise le mot « do » (entrée de base) qui est un élément de la base UNL KB. Cette base est un graphe sémantique composé de toutes les relations binaires directes entre UWs. La structure générale de l'hierarchie des UWs dans la base UNL KB<sup>2</sup>, qui est sous forme de niveaux, est illustrée dans la figure 5.

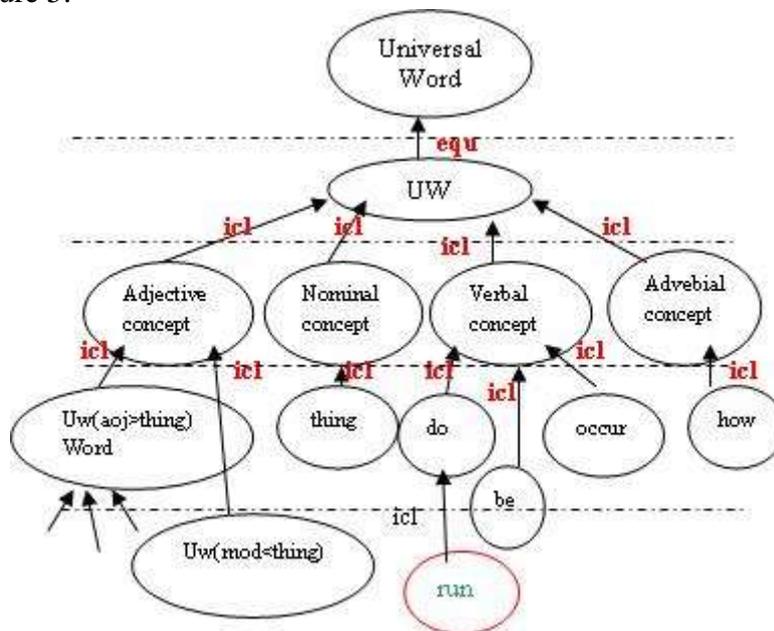


Figure 5. Structure de la base UNL.

<sup>2</sup> <http://www.unl.org>

Après avoir détaillé les principes du langage UNL, nous présentons le format d'un document multilingue UNL-HTML.

## 4. Format et document UNL-HTML

### 4. 1. Balises et syntaxe

Le document UNL (Boitet et Wang-Ju, 2002) a le format illustré dans le tableau 1.

<document UNL>	::= "[D:" <dinf> "]" { "[P:" <nombre> "]" { "[S:" <nombre> "]" <phrase> "[/S]" }... "[/P]" }... "[/D]"
<dinf>	::= <nom de document > ";" <nom de propriétaire> [ ";" <id. de document > ";" <date> ";" <E-mail > ]
<nom de document >	::= "dn=" <chaîne de caractères>
<nom de propriétaire>	::= "on=" <chaîne de caractères>
<id. De document>	::= "did=" <chaîne de caractères> /* définit par le système*/
<date>	::= "dt=" <chaîne de caractères> /* définit par le système*/
<E-mail>	::= "mid=" <chaîne de caractères> /* définit par le système*/
<phrase>	::= "{org:" <l-tag> [ "=" <code> ] }" <source de la phrase> "{/org}" "{unl" [ ":" <uinf> ] }" <expression de UNL> "{/unl}" "{ " <l-tag> [ "=" <code> ] [ ":" <sinf> "]" <phrase générée> "{/" <l-tag> }"
<l-tag>	::= "ab"   "cn"   "de"   "el"   "es"   "fr"   "id"   "hd"   "it"   "jp"   "lv"   "mg"   "pg"   "ru"   "sh"   "th"   "am" /* symboles des langues*/
<code>	::= <chaîne de caractères>
<source de la phrase>	::= <chaîne de caractères>
<phrase générée>	::= <chaîne de caractères>
<uinf>	::= <nom de système> ";" <nom d'éditeur poste> ";" <fiabilité> [ ";" <date> ";" <E-mail> ]
<sinf>	::= <nom de système> ";" <nom d'éditeur poste> ";" <fiabilité> [ ";" <date> ";" <E-mail> ]
<nom de système>	::= "sn=" <chaîne de caractères>
<nom d'éditeur poste>	::= "pn=" <chaîne de caractères>
<fiabilité>	::= "rel=" <entier>

Tableau 1. Format du document UNL

La signification des balises utilisées est expliquée dans le tableau 2.

[D:<dnf>]	Indique le début d'un document
[/D]	Indique la fin de document
[P:<nombre>]	Indique le début d'un paragraphe
[/P]	Indique la fin d'un paragraphe
[S:<nombre>]	Indique le début d'une phrase et son numéro
[/S]	Indique la fin d'une phrase
{org:<l-tag>=<code>}	Indique le début d'une phrase d'origine, le langage et le code peuvent être omis.
{/org}	Indique la fin d'une phrase d'origine
{unl:<unf>}	Indique le début d'une expression d'UNL associée à une phrase ou information nécessaire, ":<unf>" peut être omis.
{/unl}	Indique la fin d'une expression d'UNL d'une phrase.

Tableau 2. Signification des balises

## 4. 2. Illustration

```

<HTML><HEAD><TITLE>
Exemple EL/UNL
</TITLE></HEAD><BODY>
[D:dn = Séminaire CE ISIC sur UNL, on = UNL francais, mid = premier.auteur@here.com]
[P]
[S:1]
(1) {org:el} I ran in the park yesterday. {/org}
(2){UNL}
agt(run(icl>do).@entry.@past, i(icl>person))
plc(run(icl>do).@entry.@past, park(icl>lac).@def)
tim(run(icl>do).entry.@past, yesterday)
{/UNL}
(3){de dtime = 20020130-2035, deco = man}
Ich life gestern im park. {/de}
(4){es dtime = 20040130-2031, deco = UNL-SP}
Yo corri ayer en el parquet. {/es}
(5){fr dtime = 20040131-0805, deco = UNL-FR}
j'ai couru dans le parc hier. {/fr}
(6){am dtime = 20040131-0900, deco = UNL-AM}
ⵓⵔⵏ ⵉⵔⵏ ⵉⵏ ⵉⵏ ⵉⵏ ⵉⵏ . {/am} [S]
[S:2]
.... [S] [P] .... [D]
</BODY></HTML>

```

Explication :

1. Texte d'origine
2. Traduction en UNL
3. Traduction allemande
4. Traduction espagnole
5. Traduction française
6. Traduction amazighe

Les balises permettent à l'utilisateur de choisir ce qu'il désire visualiser. Si un utilisateur ne souhaite voir que la langue source et la traduction Amazighe, l'ordinateur ne rendra visible que les textes encadrés par les balises {org:el}{/org} et {Am dtime}{/Am}, ce qui donnera « I ran in the park yesterday ».

ⵓⵔⵏ ⵉⵔⵏ ⵉⵏ ⵉⵏ ⵉⵏ ⵉⵏ .

## 5. Enconvertisseur et déconvertisseur UNL-Amazighe

Chaque équipe de l'UNL travaille sur le déconvertisseur et enconvertisseur de sa propre langue, en augmentant le nombre d'UWs et en précisant les relations entre les UWs. C'est en augmentant le nombre de phrases analysées que les progrès de la traduction automatique se réalisent. Toutes les

UWs sont regroupées dans une base de données, ainsi que les relations possibles entre les UWs. La base de données d'UWs sera donc complète quand tous les concepts mondiaux seront entrés dans cette base, c'est-à-dire jamais... En effet, la langue évoluant sans cesse, de nouveaux concepts apparaissent tous les jours. Le but de l'UNL n'est pas de réaliser une traduction parfaite mais compréhensible, avec le moins d'ambiguïté; le réel danger d'une traduction étant l'ambiguïté, mais non la faute grammaticale ou d'orthographe.

## 5. 1. Enconvertisseur Amazighe-UNL

L'enconversion n'est pas (si on veut de la qualité pour du tout venant) une analyse classique. C'est une méthode de fabrication de graphes UNL qui suppose une bonne part d'interaction, avec plusieurs possibilités :

- analyse classique multiple suivi d'une désambiguïsation interactive en langue source
- entrée sous langage contrôlé,
- encore plus séduisant (et encore pas clair, au niveau recherche pour l'instant), entrée directe via une interface graphique reliée à la base lexicale (dictionnaire) et à la base de connaissances.

En effet, c'est un analyseur indépendant de langues, qui fournit une structure pour une analyse morphologique, syntaxique et sémantique simultanée en utilisant un ensemble de règles universelles et le dictionnaire des UW. Le processus d'un enconvertisseur amazighe futur est le suivant : chaque texte d'entrée est scanné de gauche à droite. Par la suite, chaque morphème correspondant aux initiales de texte est extrait de dictionnaire des UWs et il devient morphème candidat. Les règles universelles (grammaticales) seront appliquées sur ces morphèmes selon l'ordre de priorité pour construire l'arbre syntaxique et le réseau sémantique de la phrase. Le résultat de ce processus est un réseau sémantique exprimé dans le format UNL. Le schéma de la figure 6 illustre la structure d'un enconvertisseur Amazighe-UNL.

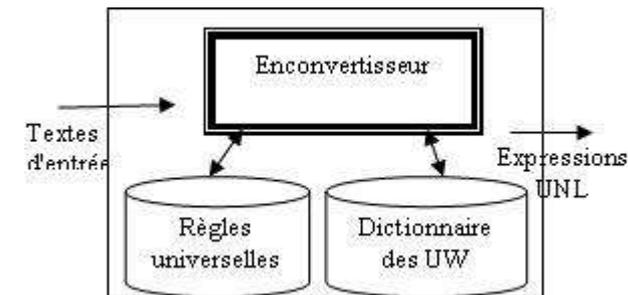


Figure 6. Structure d'enconvertisseur Amazighe

D'une manière explicite, il faut d'abord définir le corpus Amazighe dans le dictionnaire des UW en s'inspirant des anciens travaux (Achab, 96) (Achab, 98) (Arnaiz-Villena, 2000) (Haddachi, 2000). Il faut ensuite intégrer les règles grammaticales et sémantiques de l'Amazighe dans la base des règles universelles.

## 5. 2. déconvertisseur UNL-Amazighe

Le déconvertisseur est un générateur indépendant de langues naturelles qui fournit une structure pour une génération morphologique et syntaxique. Il peut déconvertir les expressions UNL à une variété de langues naturelles, en utilisant un certain nombre de données linguistiques telles que le dictionnaire, les règles grammaticales et le dictionnaire co-occurrence de chaque langue. Choses qui restent à concevoir et à réaliser par les chercheurs de la langue Amazighe.

Une question se pose : comment UNL peut être utilisé via l'Internet ? Pour répondre à cette question, on expliquera le processus dans la section suivante qui détaille le système UNL.

## 6. Système UNL

Le système UNL consiste en un ensemble de serveurs, d'éditeurs et de visionneurs UNL-langues naturelles. Prenons l'exemple d'un éditeur Amazighe et d'un visualiseur français comme illustré dans la figure 7.

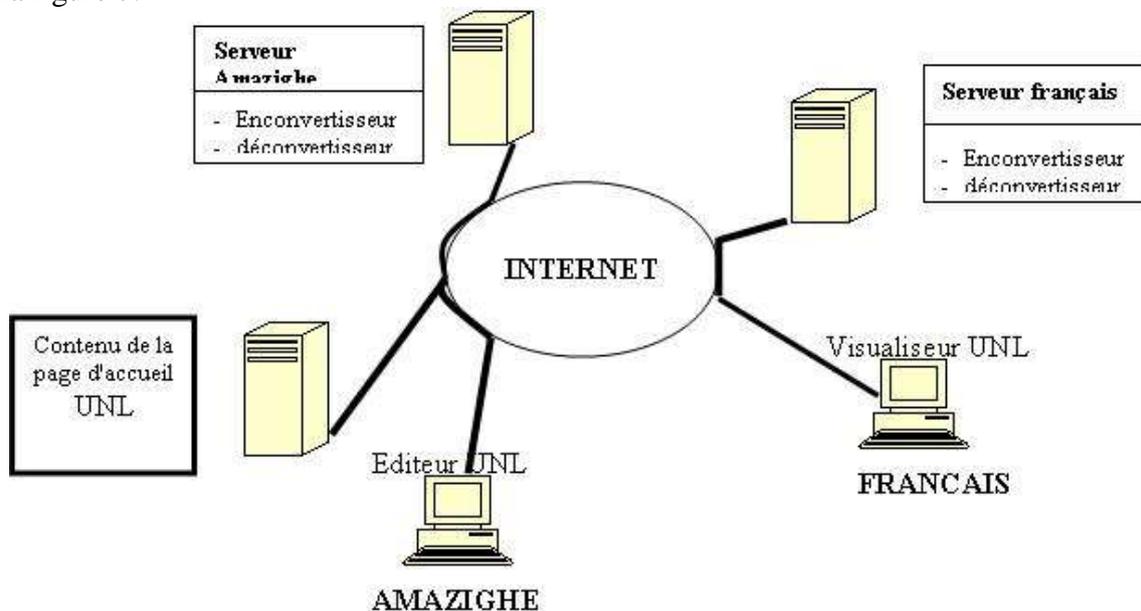


Figure 7. Structure de système UNL

Lorsqu'on développe une page Web en Amazighe, l'éditeur UNL reconnaît le contenu Amazighe et envoie le contenu au serveur Amazighe pour enconvertir le texte. Lorsque le texte est enconverti en UNL, le serveur Amazighe envoie le résultat à l'éditeur UNL. Les développeurs de page Web peuvent ensuite intégrer UNL dans leur page web.

Lorsque le visualiseur français essaie de lire cette page Web, il reconnaît le format UNL-HTML et envoie le contenu au serveur Français pour une déconverssion en français qui se charge d'envoyer par la suite le résultat au visualiseur français.

Le serveur Amazighe contiendra deux fichiers. Le premier est un dictionnaire qui liste la correspondance entre les UW et les mots de l'Amazighe. Le deuxième contiendra les règles grammaticales de l'Amazighe.

## 7. Conclusion

Le meilleur moyen de promouvoir la culture et la langue Amazighe a été d'adapter les techniques d'informations et de communication à l'apprentissage de la langue. Informatiser l'Amazighe demande un effort de plusieurs chercheurs et l'utilisation des leviers permettant d'obtenir rapidement des logiciels de qualité. Dans ce contexte, les documents multilingues resteront un moyen de diffusion du savoir Amazighe afin de valoriser cette culture sur le Web et dans le monde informatisé. En effet, il nécessaire de développer la plate forme logicielle pour intégrer l'Amazighe dans les documents multilingues. Le meilleur moyen est de développer un serveur Amazighe qui intègre l'enconvertisseur et le deconvertisseur UNL puisque l'UNL est un système de traduction automatique universel soutenu par la communauté internationale.

- Al Assimi A.-B. (2000), "Gestion de l'évolution non centralisée de documents parallèles multilingues". *Nouvelle thèse*, UJF, Grenoble, 31/10/00, 200 p.
- Al Assimi A.-B., et Boitet C. (2001) "Management of Non-Centralized Evolution of Parallel Multilingual Documents". Proc. of Internationalization Track, 10th International World Wide Web Conference, Hong Kong, May 1-5, 2001, 7 p.
- Blanc E. (2001) "From graph to tree : Processing UNL graph using an existing MT system". *Proc. of First UNL Open Conference - Building Global Knowledge with UNL*, Suzhou, China, 18-20 Nov. 2001, UNDL (Geneva), 6 p.
- Boguslavsky I. et al. (2000) "Creating a Universal Networking Language Module within an Advanced NLP System". *Proc. of COLING-2000*, Saarbrücken, 31/7—3/8/2000, ACL & Morgan Kaufmann, 1/2, pp. 83-89.
- BOITET, C, et Wang-Ju, T. (2002) " La coédiction langue↔UNL pour partager la révision entre les langues d'un document multilingue : un concept unificateur", *Traitement Automatique des langues Naturelles TALN'2*, Nancy, 24-27/06/2002 P. 275-286.
- Rachidi, A. et Mammass, D. (2005a) " Informatisation de La Langue Amazighe : Méthodes et Mises En Œuvre", *Setit 2005 ' 3rd International Conference: Sciences Of Electronic Technologies Of Information And Telecommunications ' March 27-31, 2005 – Tunisia*.
- Rachidi, A. Mammass, D. (2005b) "Vers Un Système D'écriture Informatique Amazighe : Méthodes Et Développements", *Recital 2005, Dourdan, 6-10 Juin 2005*
- Rachidi, A. et Mammass, A. (2005c) "Methods and developments in the creation of a computerised Amazigh script", *European Language Resources Association letter's, the ELRA Newsletter, Vol.10 N°1, January-March 2005*.
- Sérasset G., et Boitet C. (1999) "UNL-French deconversion as transfer & generation from an interlingua with possible quality enhancement through offline human interaction". *Proc. of MT Summit VII*, Singapore, 13-17 September 1999, Asia Pacific Ass. for MT, pp. 220—228.
- Sérasset G., et Boitet C. (2000) "On UNL as the future "html of the linguistic content" & the reuse of existing NLP components in UNL-related applications with the example of a UNL-French deconverter". *Proc. of COLING-2000*, Saarbrücken, 31/7—3/8/2000, ACL & Morgan Kaufmann, 2/2, pp. 768—774.
- Uchida H. (1989) "ATLAS". *Proc. of MTS-II (MT Summit)*, Munich, 16-18 août 1989, pp. 152-157.
- Boukous, A. et al," L'amazighe : les défis d'une renaissance", *Nouvelle livraison de Prologues* Numéro double 27/28, 2003 ,
- Achab, R. (1996) "La néologie lexicale berbère", (1945-1995), Paris/Louvain : Ed. Peeters, collection, « M.S. – 9 – Ussun amari », , 367p., préf. de S. Chaker.
- Achab, R. (1998) " Langue berbère – Introduction à la notation usuelle en caractères latins", Paris : Ed. Hoggar,1998.
- Arnaiz-Villena, A. (2000), "Prehistoric Iberia : Genetics, Anthropology, and Linguistics", New York: Kluwer Academic/Plenum Publishers, 2000.
- Haddachi, A. (2000) " Dictionnaire de Tamazight. Parler des Ayt Merghad (Ayt Yafelman)", Salé : Imprimerie Béni Snassen,.